

SUPERVISOR'S APPROVAL

PUBLIC OPINION EXTRACTION AND VISUALIZATION FROM TWITTER SOCIAL MEDIA PLATFORM

By

AHMAD ZAKI FITRI BIN ROSLAN

2012616836

This report was prepared under the supervision of the project supervisor, Mr Muhammad Bakri Bin Che Haron. It was submitted to the Faculty of Computer and Mathematical Sciences and was accepted in partial fulfillment of the requirements for the degree of Bachelor of Computer Science (Hons)

Approved by

.....
Muhammad Bakri Bin Che Haron
Project Supervisor

JULY 30, 2015

STUDENT'S DECLARATION

I certify that this report and the project to which it refers is the product of my own work and that any idea or quotation from the work of other people, published or otherwise are fully acknowledged in accordance with the standard referring practices of the discipline.

.....
AHMAD ZAKI FITRI BIN ROSLAN
2012616836

JULY 30, 2015

ABSTRACT

Sentiment analysis in social network has emerged as an alternative and effective method to study human behavioral through their social interaction in the social media. In the current scenario, the number of social media user is growing rapidly, results in the rise of micro blog's popularity among internet user. In this project, Twitter, a micro blogging website has been used to gather public opinion on selected trending topics. However, gathering opinion from Twitter that have large amount of tweets daily is not a simple task as the opinion could be hidden in the pile of post. So, it is hard for human readers to extract the relevant opinions, summarize and organize them in a usable format. This project purpose a combination of techniques to assist in extraction and visualization process. The extraction process of the tweets will be done by using the REST API provided by Twitter. The extracted tweets will undergo a cleaning process to remove unrelated words and naive Bayes classifiers will be used to identify either the tweet is positive or negative. The classified tweet and the frequency of the unique word will be taken for the visualization process. The visualization process is done by using d3 word cloud. The project will implement the ADDIE model for the framework as it is suitable for data visualization. As a significance of the study, this project will give a review about peoples' opinions regarding the hot topics or latest news that been discussed in social media which Twitter is focused as the platform. Further research might explore the monitoring of sentiments in Twitter in a much detailed way. Another form of sentiment in sentences such as sarcasm, factual, or mood expression can be recognized by the system. The system should also able to reduce noisy data as much as it can in order to have a more accurate sentiment analyzer.

Keyword: Sentiment analyzer, bayes classifier, word cloud.

TABLE OF CONTENTS

CONTENTS	PAGE
SUPERVISOR'S APPROVAL	ii
STUDENT'S DECLARATION	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	ix
LIST OF TABLES	xi
CHAPTER 1	1
1.0 Introduction	1
1.1 Background of Study	1
1.2 Problem Statement	2
1.3 Research Purpose	3
1.4 Research Objective	3
1.5 Project Scope	3
1.6 Significance of the project	3
CHAPTER 2	4
2.0 Introduction	4
2.1 Social Networking	4
2.1.1 History	4
2.1.2 Social Networking Type	6
2.1.3 Twitter as social networking platform	8
2.2 Data Extraction	9
2.2.1 Data Extraction Techniques	10
2.3 Text Classification	12
2.3.1 Text classification techniques	12

2.3.2	Naïve Bayes Classifier	13
2.3.3	Bayes Techniques	14
2.3.4	Current Methods of Social Network Data Analysis	15
2.4	Data Visualization	16
2.4.1	Type of data visualization	16
2.4.2	Word Cloud	21
2.4.3	Type of word cloud	22
2.5	Conclusion	23
CHAPTER 3		24
3.0	Project Methodology	24
3.1	Project Framework	24
3.2	Detail Research Methodology	25
3.2.1	Preliminary Study	27
3.2.2	Prototype Design	28
3.2.3	Development	42
3.2.4	Testing	43
3.2.5	Result Analysis	44
3.3	Software and Hardware Requirement	45
CHAPTER 4		46
4.0	Introduction	46
4.1	Prototype Main Interface	46
4.2	Testing	47
4.2.1	Functionality Testing	47
4.2.2	Correctness Testing	49
4.2.3	System Testing	51
CHAPTER 5		69
5.0	Introduction	69
5.1	Conclusion	69
5.2	Limitations	69
5.3	Future Recommendation	70
5.4	Summary	70